Title

A New Change-of-Contents False Belief Test: Children and Chimpanzees Compared

Journal
International Journal of Comparative Psychology, 23(2)

ISSN
0889-3675

Authors
Krachun, Carla
Carpenter, Canada Malinda
Call, Josep
et al.

Publication Date
2010

DOI
10.46867/ijcp.2010.23.02.03

International Journal of Comparative Psychology, 2010, 23, 145-165.

A New Change-of-Contents False Belief Test: Children and Chimpanzees Compared

Carla Krachun
Max Planck Institute for Evolutionary Anthropology, Germany
Carleton University, Canada

Malinda Carpenter, Josep Call, and Michael Tomasello
Max Planck Institute for Evolutionary Anthropology, Germany

In the handful of existing comparative false belief studies, chimpanzees have consistently failed tests that 5- to 6-year-old children have passed. However, those tests were either explicitly cooperative-communicative or competitive, both of which are problematic for different reasons. We therefore devised a new *change-of-contents* false belief test for children and chimpanzees that did not include these problematic elements. Nevertheless, chimpanzees showed no evidence of understanding false beliefs (consistent with past research, however, children showed a clear improvement in test performance from 3.5 to 4.5 years of age). Our results suggest that the cooperative-communicative or competitive nature of previous false belief tests was not solely responsible for chimpanzees' failure. That chimpanzees have now also failed a more socially neutral test supports the conclusion that chimpanzees may simply not recognize false belief states in others. Additionally, our test departs from a near exclusive reliance on the change-of-location paradigm in false belief research. It therefore expands the repertoire of methods available for testing false belief understanding with minimally verbal and nonverbal procedures.

Being able to recognize when others have a false belief about some state of affairs is an important development in children's growing understanding of the mind, one that many would argue does not occur in any other species. It demonstrates, beyond a doubt, children's recognition that the mind does not simply reflect reality but represents it, and that in doing so it can sometimes get things wrong. Among the hundreds of studies aimed at understanding when and how false belief understanding emerges in human ontogeny, two paradigms have been so widely used that they have become known as the standard verbal false belief tests. The first is the *Maxi-chocolate* or *Sally-Anne* test, in which children are asked to predict where a character will first look for an object that was moved to a new location in her absence (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). The correct answer, of course, is that she will look for the object in its original location, where she falsely believes it to still be. In the other standard test, called the *Smarties* test (Perner, Leekam, & Wimmer, 1987), children first discover that a candy box really contains some other item (e.g., pencils), and they must predict what someone else will think is inside the box before it is opened. Children

A number of people provided invaluable assistance with testing and coding. For help with the children's study, we thank Jana Jurkat, Juliane Dauksch, Susanne Grassmann and Kristin Liebal; for help with the chimpanzee study, we thank Keith Jensen, Alexandra Gross, Ingrid Kaiser and Stefanie Keupp. We would also like to thank two anonymous reviewers for their helpful comments on an earlier version of this paper. Correspondence concerning this article should be addressed to Malinda Carpenter, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, D-04103, Germany. (carpenter@eva.mpg.de).

who answer ‘candy’ demonstrate their understanding that a naïve viewer of the box will falsely believe it to contain the item depicted on its label.

Typically, children begin to pass these standard tests of false belief understanding between about 4 and 5 years of age (see a meta-analysis by Wellman, Cross, & Watson, 2001), although younger children have also passed some variations on the standard paradigms. For example, Sullivan and Winner (1993) found that 3-year-olds who failed a Smarties test could nevertheless pass it when it was turned into a game of trickery with children actively participating in the deception by switching the contents of the box. In two other studies, some 3-year-olds were capable of intentionally deceiving someone about the location of a treasure even when they had to figure out how to do so themselves. For example, they wiped away footprints leading to the actual location of a hidden treasure and added footprints leading to another, incorrect location (Chandler, Fritz, & Hala, 1989; Hala, Chandler, & Fritz, 1991; but see also Sodian, Taylor, Harris, & Perner, 1991). Younger children have also performed better in false belief tests involving active helping responses (Buttelmann, Carpenter, & Tomasello, 2009b; Carpenter, Call, & Tomasello, 2002; Matsui & Miura, 2008).

A number of researchers have addressed the possibility that younger children fail standard tasks not for lack of understanding, but because their grasp of language is not developed enough to deal with the tasks’ high linguistic demands. Supporting this, 3-year-olds performed better in a false belief test by Carlson, Wong, Lemke, and Cosser (2005) that focused on gestures rather than verbal responses. Also, Clements and Perner (1994) observed that while 3-year-olds made the incorrect choice in a change-of-location test, they did look at the correct choice, suggesting they had some understanding of false beliefs that was not reflected in their active choice responses. A number of recent studies have reported similar looking responses suggestive of false belief understanding in children as young as just 1–2 years of age (Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007).

The question of whether or not false belief understanding is a uniquely human ability or is also present in our closest living evolutionary relatives has been the subject of relatively little empirical investigation. In recent years, just a handful of studies have compared the performance of children and apes on equivalent tasks, all of them using nonverbal adaptations of the change-of-location paradigm. Call and Tomasello (1999), for example, devised a change-of-location test for 4- to 5-year-old children, orangutans, and chimpanzees. During training trials a ‘Communicator’ watched a ‘Hider’ stash a reward in one of two boxes (participants themselves could not see where the Hider put the reward). The Communicator next marked the correct box for participants by briefly placing a wooden block on top of it, and participants were then allowed to choose a box. In the subsequent false belief test, after the hiding event the Communicator immediately left the area. While she was gone, the Hider transposed the location of the boxes, causing the Communicator to return and mark the incorrect box. Five-year-old children demonstrated false belief understanding by choosing the unmarked box in these trials, but 4-year-old children and apes did not. All groups, however, passed a true belief control test in which the Communicator witnessed the location switch and thus marked the correct box, indicating that even the younger children and apes were capable of performing well when there were no false beliefs involved. In a similar task by O'Connell and Dunbar (2003), 6-year-old children did well in both the false belief trials and the true belief control trials, while younger children only did well in the latter. The four chimpanzee participants passed the false belief test trials but, importantly, they failed the true belief control trials, thereby casting serious doubt on the false belief results.

In short, neither of these studies provided any convincing evidence that apes were capable of recognizing others' false belief states. However, these tests may have been problematic for apes because they involved a cooperative experimenter who voluntarily communicated the location of the hidden food. Chimpanzees normally compete for food (Hare, 2001), and they have been shown to have difficulties understanding communicative intentions (Hare & Tomasello, 2004). Two recent comparative false belief studies therefore used competitive paradigms to make the tests more species-relevant for the ape participants. Krachun, Carpenter, Call, and Tomasello (2009) tested children, chimpanzees, and bonobos in a task similar to Call and Tomasello's (1999) change-of-location test described above. The crucial difference, however, was that rather than having a Communicator mark the container for participants, they had a Competitor reach towards a container (the wrong one in false belief trials) with the clear intent of taking the reward for herself. Apes did not choose the correct container in this test, while 5-year-old children did. Chimpanzees also failed (and 6-year-old children passed) a false belief test by Kaminski, Call, and Tomasello (2008, Study 2) in which participants competed against conspecifics rather than humans. These studies therefore provided little evidence that using a competitive paradigm helps apes to pass false belief tests.

Nevertheless, it may be unfair to conclude on the basis of this small collection of studies that chimpanzees do not possess any understanding of false beliefs, because the previous work is limited in two important respects. First, the tests used have been either explicitly cooperative-communicative or competitive, both of which may be problematic for different reasons. As mentioned above, having someone willingly communicate the location of desired food may be too unnatural for apes and thus difficult for them to comprehend. On the other hand, competitive contexts could potentially trigger evolutionarily primed, prepotent responses that could interfere with false belief performance (see Keenan & Ellis, 2003). That is, when cognitive tasks involve elements similar to situations crucial for survival in a species' evolutionary history (such as competing for food), they could activate response patterns that evolved for dealing quickly and effectively with those situations, thus making careful, thoughtful responses difficult. This effect could be especially disruptive in computationally complex tasks, such as monitoring which events in an unfolding sequence others have and have not witnessed and then combining these separate pieces of information to judge the resulting belief states. Indeed, Keenan and Ellis (2003) found that 3- to 4-year-old children's performance deteriorated when a false belief test was framed in terms of predator avoidance, another situation likely to activate automatic, evolved mechanisms. In nonhuman primates, Washburn, Hopkins, and Rumbaugh (1990) observed deleterious effects of competition on response accuracy in rhesus monkeys performing a video task. Furthermore, while chimpanzees have performed well in knowledge-ignorance tests in competitive contexts (Hare, Call, & Tomasello, 2001; Kaminski et al., 2008, Study 1), such tests may be computationally simpler than false belief tests (Hogrefe, Wimmer, & Perner, 1986). Thus, the competitive nature of some previous false belief tests, although intended to aid performance, may have inadvertently hindered it. Apes may have performed poorly because competing is *too* natural for them.

The second limitation is that all of the comparative studies described above have used the same general procedure: some nonverbal adaptation of the standard verbal change-of-location paradigm. A greater variety of tests suitable for use with both young children and apes is needed. In studies with children, inconsistencies in performance across different false belief tests—even across standard tests—have sometimes been found (e.g., Charman & Campbell, 1997; Holmes, Black, & Miller, 1996; Naito, 2003). Such inconsistencies are likely due to individual differences in false belief competence across different types of procedures. Being competent in one type of false belief test is clearly no guarantee that one will be competent in another. It is thus possible that while apes may not be capable of passing false belief tests that rely on the change-of-location paradigm, they could potentially be capable of passing false belief tests that use other paradigms.

In this paper, we address the limitations noted above with a new *change-of-contents* false belief test for children and chimpanzees. The version developed for children is necessarily somewhat different from the version developed for apes, mostly in terms of the materials used. In both cases, however, in order to succeed participants must correctly predict an experimenter's actions by recognizing her false belief about the hidden contents of a container. Unlike previous comparative false belief tests, our test is neither explicitly cooperative-communicative nor competitive but instead takes place in a more neutral social context. The experimenter tries to hide a reward such that participants will be able to find it, but she does not explicitly indicate the location of the reward for participants. By avoiding this communicative feature of some previous tests, as well as the competitive element of other tests, we have attempted to create a more level playing field between chimpanzees and children, increasing the validity of the cross-species comparison. In addition, our new false belief test departs from a previous over-reliance on variations of the change-of-location test by focusing less on location and more on contents, as in the standard verbal Smarties test. In our test, a false belief is instilled in the experimenter by switching the *contents* rather than the location of a container in her absence. We describe the children's version of the test first below, followed by the chimpanzees' version.

Study 1: Children

Method

Participants

Participants included 24 children in each of three age groups: 3.5 - 4 years old (M = 45.4 months, SD = 1.6 months); 4 - 4.5 years old (M = 50.6 months, SD = 1.7 months); and 4.5 - 5 years old (M = 57.3 months, SD = 1.8 months). Children were recruited from daycare centers by letters sent to their parents.

Design and materials

Children received two false belief trials and two true belief trials, blocked and counterbalanced for order. A sticker was used as the reward for one of each type of trial and a rubber stamp for the other, in counterbalanced order. To conceal the reward, we used an opaque plastic Easter egg that could be opened by pulling its two halves apart. The apparatus was a box measuring 54 x 36 x 38 cm with two 7-cm diameter holes on the top opening into two chutes (see Figure 1). One chute led to an opening at the bottom right side of the box and the other chute led to an opening at the bottom left. Curtains covered these openings, and a sticker attached to the box above the right curtain designated it the ‘sticker side’ while a rubber stamp attached above the left curtain designated it the ‘stamp side’. Two small boxes measuring 21.5 x 15 x 7 cm were attached to the bottom of the large box, one on each side, to obscure children’s view of the openings. During test trials and some training trials (see below), a third small box attached to the top of the large box blocked children’s view so that they could not see which hole the experimenter put the egg into. The chutes were padded so that children could not hear the egg moving through them.

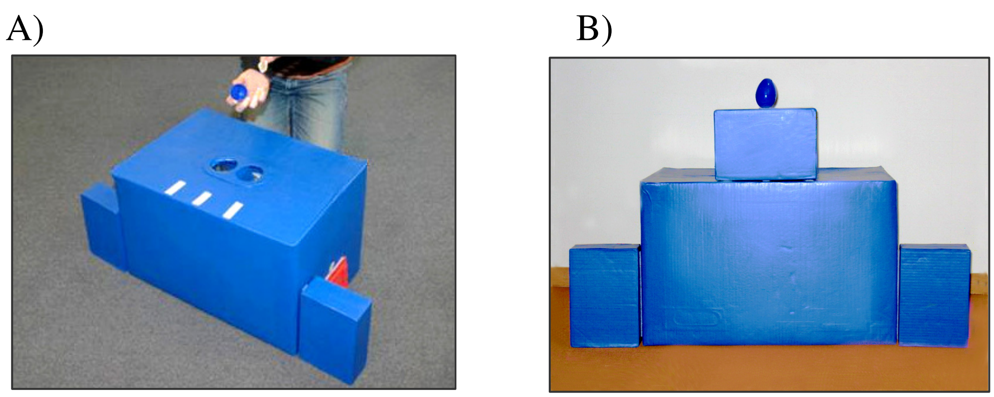

Figure 1. Apparatus for Study 1: A) at the beginning of training, with one of the side openings just visible behind the small box on the right; and B) as seen from the child’s perspective during test trials, with children’s view of the holes blocked.

In addition to the new change-of-contents test, children participated in two standard verbal false belief tests: the Sally-Anne test (Baron-Cohen et al., 1985) and the Smarties test (Perner et al., 1987), including the customary control questions. Half the children received both tests before the change-of-contents test and half received them afterwards, and the two tests were counterbalanced for order. For the Sally-Anne test, two different dolls, a covered basket, an opaque plastic box, and a small ball were used. For the Smarties test, a Smarties candy box with a candle inside was used (or an egg carton with a toy car inside, if the child was not familiar with Smarties).

Procedure

Children were tested in their daycare center by two female experimenters: the 'Baiter' was responsible for putting the reward into the egg and for putting the egg into the box, and the 'Switcher' was responsible for switching the contents of the egg during test trials. Before the test, children received training to familiarize them with the apparatus and the rules of the game.

Training procedure. Children were first taught that when there was a sticker inside the egg, the Baiter always put that egg into the hole on the right (from their perspective) in the top of the box, causing it to travel down a chute to the bottom right side of the box. They were likewise taught that when there was a stamp inside the egg, the Baiter put that egg into the left hole in the top of box, causing it to travel down to the bottom left side of the box. The procedure by which children were taught this was as follows:

First, children were shown the apparatus without the small box on top, so that the two holes there were visible (see Figure 1A). They were encouraged to reach down each chute so that they understood the mechanics of the box. The Baiter sat behind the apparatus, facing the child and the Switcher. As the child observed, the Baiter held up a sticker, put it into the egg, closed the egg, and put the egg into the hole on the child's right, explaining that stickers always go on that side. The Switcher then showed the child how the egg was retrievable in the opening behind the curtain on the bottom right side of the box, and she pointed out the sticker reminder above the curtain. The Switcher helped the child get the egg, remove the sticker, and stick it onto a piece of paper, which the child could keep. This procedure was repeated on the left side with a rubber stamp. Children were told they could only have the reward if they went to the correct side of the box on their first try.

Next, the small box was placed on top of the apparatus so that children could not see which hole the Baiter put the egg into (see Figure 1B). The Baiter then carried out a number of trials in which she put either a sticker or a stamp (determined randomly) into the egg as the child observed. With the child's view of the holes blocked, she then put the egg into the appropriate hole in the top of the box. Thus, in these trials, children knew which type of reward was in the egg but they could not see which hole the Baiter put the egg into. They needed to use the rule they had learned in the earlier phase of training: stickers were put into the right hole and stamps were put into the left. At the end of each trial, the Switcher encouraged children to retrieve the egg. When children could correctly retrieve, on their first try, three stickers and three stamps consecutively (i.e., after six correct trials), they proceeded to the test trials. Children typically reached this criterion within six to nine trials.

Test trials. For the false belief trials, the Baiter began as before by putting either a sticker or stamp into the egg as the child watched and then closing the egg. This time, however, she set the egg down on top of the apparatus, saying that she had to leave but would be right back. When the Baiter had left the room and shut the door the Switcher said, "Let's play a trick on [Baiter's name]!" She sneakily removed the reward from the egg and replaced it with the other type of reward, then closed the egg and replaced it on top of the apparatus. The Baiter then returned, picked up the egg, and (with her hands hidden from view by the small box) put it into a hole in the top of the apparatus. She then turned away and began writing, so that she did not witness the child's response or the actual contents of the egg. (This was done so that children would not think the Baiter might anticipate a switch in future trials). The Switcher then encouraged children to retrieve the egg. If children understood that the Baiter had a false belief about the contents of the egg because she did not see the switch, they should go to the side of the apparatus corresponding to the *original* contents of the egg.

The two true belief trials were identical to the false belief trials except that the Baiter witnessed the switch and it was not sneaky. Thus, the Baiter put one of the rewards into the egg, closed the egg and put it down, and left the room for 30 seconds, the same average amount of time as in the false belief condition. During this time, the Switcher and the child examined and discussed the stickers and stamps the child had accumulated. When the Baiter returned, she and the child observed as the Switcher switched the contents of the egg. The Baiter then put the egg into a hole in the top of the apparatus (behind the small box) and turned away to write while children responded. If children understood that the Baiter had a true belief about the contents of the egg, because she had in this case witnessed the switch, they should go to the side of the apparatus corresponding to the new, *current* contents of the egg.

When children responded correctly and retrieved the egg on their first attempt they could keep the reward. When they responded incorrectly by going to the wrong side of the apparatus, the Switcher showed them that the egg was in the other opening and put it away without the Baiter seeing. Children received a score of 0–2 for each condition, corresponding to their number of correct trials. To minimize confusion for children, in between the two false belief and two true belief trials we administered at least two normal training trials with no switch (i.e., until children responded correctly at least twice in a row).

Coding and analyses

Children's choice was defined as moving to one side of the apparatus and lifting the curtain to reach inside the opening. It was unambiguous and therefore coded live. To determine inter-rater reliability, an independent coder blind to condition and naïve to the hypotheses of the study coded 21% of the videotapes for each age group, and there was 100% agreement. All *p* values reported below are two-tailed unless otherwise noted.

Results and Discussion

Children's performance was not significantly affected by the order of (1) the Sally-Anne and Smarties tests, (2) those two tests and the change-of-contents test, or (3) the false belief and true belief trials within the change-of-contents test (Fisher tests: all *p*s > 0.16). We therefore collapsed across these orders for all analyses.

Change-of-contents test

Figure 2 presents the mean proportion of trials children got correct in each condition at each age for the change-of-contents test. All groups performed better in true belief trials compared to false belief trials (Wilcoxon tests: 3.5 - 4-year-olds: $z = 3.76$, $p < 0.001$; 4 - 4.5-year-olds: $z = 2.82$, $p = 0.005$; 4.5 - 5-year-olds: $z = 3.22$, $p = 0.001$). However, both false belief performance (Kruskall-Wallis test: X^2 (2, 72) = 10.09, $p = 0.003$, one-tailed) and true belief performance (Kruskall-Wallis test: X^2 (2, 72) = 5.73, $p = 0.029$, one-tailed) improved significantly with age.

When comparing performance in the false belief condition to the chance probability of success (50%), there was a clear developmental pattern: 3.5 - 4-year-olds were worse than chance (Wilcoxon test: $z = 2.18$, $p = 0.029$), 4 - 4.5-year-olds were at chance (Wilcoxon test: $z = 0.73$, $p = 0.47$), and 4.5 - 5-year-olds were above chance (Wilcoxon test: $z = 2.32$, $p = 0.020$). In contrast, all age groups performed above chance levels in the true belief test (Wilcoxon tests: $z > 4.35$, $p < 0.001$ in all cases). This is in line with past research using the change-of-location paradigm, in which dramatic shifts in false belief performance between 3 and 5 years of age have often been observed (Wellman et al., 2001). Furthermore, it seems that children found our false belief test somewhat easier than previous ones created for use with both children and apes. Comparable comparative studies that have tested children developmentally have reported success for children only at 5 years old (Call & Tomasello, 1999) or 6 years old (O'Connell & Dunbar, 2003), with younger children performing poorly.

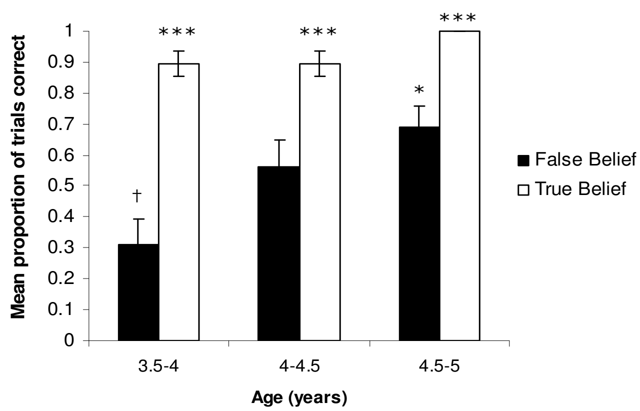

Figure 2. Mean (±*SE*) proportion of trials correct in the false belief and true belief tests in Study 1 as a function of age (within-subjects, N = 24 for each bar). $^*p < 0.05$, $^{***}p < 0.001$ (above chance); †$p < 0.05$ (below chance).

Comparisons among tests

To compare children's performance on the two standard verbal tests and our new change-of-contents test on an equal footing, we used only the first false belief trial for each child from the change-of-contents test. As usual, children only passed the standard verbal tests if they also answered the customary control questions correctly (see Baron-Cohen et al., 1985; Perner et al., 1987). For the change-of-contents test, we used the first true belief trial as a control question. Thus, for this analysis, children passed the change-of-contents test only if they got both the first false belief trial and the first true belief trial correct.

Figure 3 presents the number of children who responded correctly to both the test and control questions as a function of age. In all three age groups, there were significant differences among tests (Cochran's $Q \geq 8.44$, $df = 2$, $p \leq 0.015$, $N = 24$ in all cases). Pairwise comparisons revealed, however, that for the two oldest groups, the only significant differences were between the Sally-Anne test and the Smarties test (Sign test: $p = 0.013$ in both cases). In the youngest group, Sally-Anne test performance differed from performance in the change-of-contents test (Sign test: $p = 0.004$) and it approached being significantly different from Smarties test performance (Sign test: $p = 0.065$). All other differences between tests in all groups were non-significant (Sign tests: all ps ≥ 0.15). In general, then, children appeared to find the Smarties test the most difficult and the Sally-Anne test the easiest, with our change-of-contents test falling in between the two standard verbal tests.

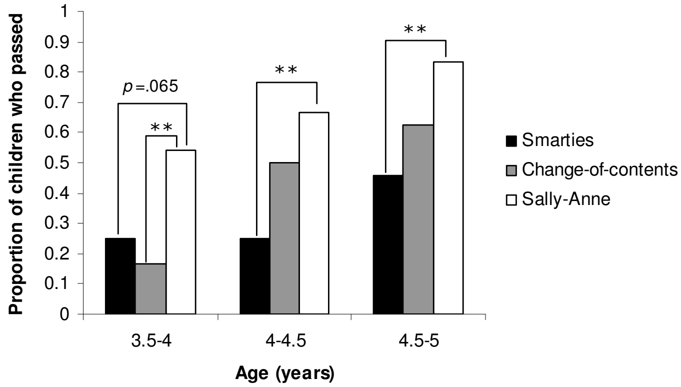

Figure 3. Proportion of children in Study 1 who responded correctly to the test and control questions in the change-of-contents false belief test (first trial only) and the two standard verbal tests as a function of age (within-subjects, $N = 24$ for each bar). $^{**}p \leq 0.01$.

Finally, we investigated relations among the tests, controlling for age. Performance on the change-of-contents test and the Sally-Anne test was positively correlated, although not very strongly (Pearson $r(69) = 0.24$, $p = 0.043$). There was no significant correlation between the Smarties test and the Sally-Anne test (Pearson $r(69) = 0.11$, $p = 0.35$) or between the Smarties test and the change-of-contents test (Pearson $r(69) = 0.18$, $p = 0.14$). Given the inconsistencies in children's performance across even well-established false belief tests (e.g., Charman & Campbell, 1997; Holmes et al., 1996; Naito, 2003) these non-correlations were not an unexpected finding.

In summary, our results for Study 1 agree well with results from previous research on preschoolers. There was a clear developmental progression from below-chance performance in our youngest group to above-chance performance in our oldest group. In the two older age groups, there were no significant differences in children's performance in our change-of-contents test as compared to both standard false belief tests. Also, individual performance in our test was positively correlated with performance in the Sally-Anne test. Having established our new change-of-contents paradigm as a valid and useful measure of false belief understanding, our next step was to adapt it for use with chimpanzees.

Study 2: Chimpanzees

As noted in the introduction, chimpanzees and other apes have performed consistently poorly in false belief studies using the change-of-location paradigm (Call & Tomasello, 1999; Kaminski et al., 2008, Study 2; Krachun, Carpenter, et al., 2009; O'Connell & Dunbar, 2003). However, those tests may have been problematic because they were either explicitly cooperative-communicative or competitive. Additionally, they all relied on variations of the change-of-location procedure, and so apes' capacity to recognize false beliefs had never before been tested outside of that paradigm. In this study, we presented chimpanzees with an adapted version of the change-of-contents test developed for children in Study 1. Thus, as in the children's study, the social context for our new false belief test was more neutral than in previous tests, and the switch leading to a false belief was a switch in the hidden contents of a container rather than in the location of a hidden reward.

Because we wanted to give chimpanzees every possible opportunity to pass our test, after the main change-of-contents test we ran two follow-up tests aimed at making it easier for them to recognize the false belief of the deceived experimenter (called the Baiter, as in Study 1). The first was the *duping* test, in which some of the chimpanzees were given personal experience being tricked regarding the contents of a container, as in the standard Smarties test (Perner et al., 1987). The second was the *in-room* test, in which the Baiter stayed in the room within view of the chimpanzees at all times but had her back turned during the switch (this manipulation improved children's performance in the study by Krachun, Carpenter, et al., 2009). Further, in addition to recording whether or not chimpanzees made the correct choice in our test, we measured their spontaneous looking behavior. Recall that several studies with children have found looking responses suggestive of false belief understanding at much younger ages than are normally associated with success in false belief tests (Clements & Perner, 1994; Onishi & Baillargeon, 2005; Southgate et al., 2007; Surian et al., 2007). Similarly, Krachun, Careptner, et al. (2009) found some evidence in one version of their test that apes might have understood more about false beliefs than their active choice responses indicated: while apes most often incorrectly chose the container the competitor was reaching for, before doing so they sometimes looked at the correct container. That they did so significantly more often in false belief trials than in true belief trials suggested they may have recognized the competitor's false belief on some level, making them hesitant about choosing the same container she did. While this result was not replicated in a second version of Krachun, Carpenter, et al.'s (2009) test, this may have been due to order effects because the two versions were not counterbalanced. We therefore thought it important to measure apes' looking responses in the current study to see if Krachun, Carpenter, et al.'s (2009) positive looking result could be replicated.

Method

Participants

Five chimpanzees housed together at the Wolfgang Köhler Primate Research Center in the Leipzig Zoo, Germany, participated in this study. The group included one male and four females from 5 to 13 years old (M = 10.2 years, SD = 3.9 years). An additional, sixth chimpanzee began testing but had to be dropped from the study for failing to achieve the training criterion, described below. Chimpanzees participated voluntarily, were not food deprived, and had water available *ad libitum*.

Design and materials

Before the test sessions, chimpanzees participated in a variable number of training sessions (12 trials per session) as needed until they reached criterion (see below). They then received a total of 20 false belief and 8 true belief trials over seven test sessions. Sessions 1–4 (main test) each included four true belief or four false belief trials (never both in the same session), with order counterbalanced across participants. Sessions 5–6 (duping test) and Session 7 (in-room test) each included four false belief trials and no true belief trials.

Rewards were grapes or banana slices. We did not use the box apparatus used with children in Study 1 because during pilot testing the chimpanzees were taking over 700 trials on average to learn how the chutes worked. To simplify the apparatus for chimpanzees, we used a set-up that they were very familiar with from previous studies: containers on a table top. Thus, the chimpanzees' apparatus consisted of a sliding table mounted below a Plexiglas window looking onto the chimpanzees' enclosure, a small yellow box with lid, and two larger, round, opaque plastic containers: one blue and one white (see Figure 4). The blue and white containers were equivalent to the two chute openings in the children's apparatus and the yellow box was equivalent to the plastic egg. Chimpanzees could choose a container by poking their fingers through holes in the bottom of the window. The location of the reward was randomized with the constraint that it could not be in the same location for more than two consecutive trials.

Note that the difference in materials used did not change the basic structure of the task. In both the children's test and the apes' test, there was a reward hidden inside a container and there were two possible places the Baiter could put that container. Participants had to understand that the Baiter would put the container in the correct place when she had a true belief about its contents and in the wrong place when she had a false belief about its contents.

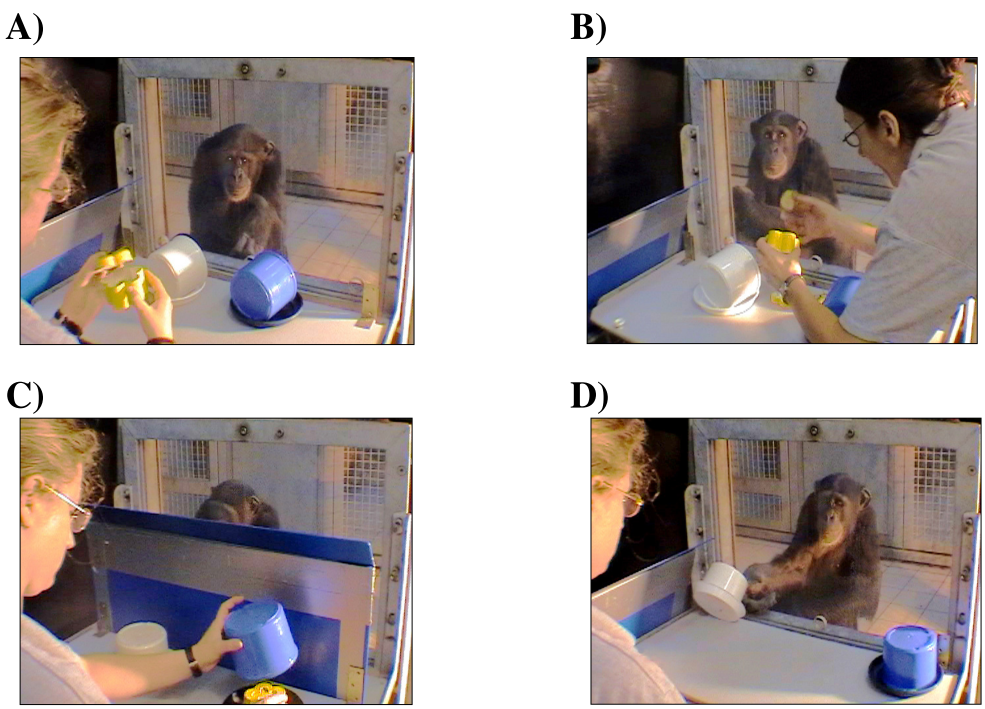

Figure 4. False belief procedure for chimpanzees in Study 2: A) the Baiter puts a grape into the box and then leaves; B) the Switcher enters and sneakily switches the grape for a banana slice; C) the Baiter returns and, because she believes it still has a grape inside, hides the box in the 'grapes' container; and D) the chimpanzee incorrectly chooses the 'banana' container.

Procedure

Chimpanzees were tested in a familiar enclosure by two female experimenters, the 'Baiter' and the 'Switcher.' To engage the chimpanzees' attention and keep them motivated, the experimenters spoke and gestured to the chimpanzees throughout the task as they would during their normal, everyday interactions with the animals.

Training procedure. Prior to testing, chimpanzees had to learn the rule that grapes went into the blue container on their left and bananas went into the white container on their right. All sessions began with two warm-up trials in which the Baiter placed, in random order, a grape (in one trial) and a banana slice (in the other trial) into the appropriate container in plain view and then let the chimpanzee choose. After this warm-up, the training trials were run. Training proceeded in three stages, with chimpanzees advancing from one stage to the next when they achieved at least 10/12 trials correct in two consecutive sessions (i.e., significantly better than chance at $p < 0.001$).

In stage 1, the Baiter showed the chimpanzee that she had either a grape or a slice of banana, and she then positioned an opaque occluder to block the chimpanzee's view of the blue and white containers. She hid the food in the appropriate container (in the blue container if it was a grape, in the white container if it was a banana), then removed the occluder and let the chimpanzee choose a container. When chimpanzees chose correctly in at least 10/12 trials for two consecutive sessions, they advanced to training stage 2.

In stage 2, the Baiter introduced the yellow box. She showed the chimpanzee the grape or banana slice, placed it into the yellow box, and closed the lid. She then positioned the occluder, hid the box in the appropriate container, removed the occluder, and let the chimpanzee choose. When chimpanzees chose correctly in at least 10/12 trials for two consecutive sessions, they advanced to training stage 3.

In stage 3, the Baiter showed the chimpanzee the grape or banana slice, placed it into the yellow box, then put the box down onto the table and left the room, closing the door behind her. She stayed outside the room for 10 seconds, approximately the same amount of time she would be gone in the true belief and false belief trials. She then returned and completed the trial as before. The second experimenter, the Switcher, also sat in the room during this last stage of training but did not participate in the procedure. She did this so that the chimpanzees would become used to her presence before the experimental trials. When chimpanzees chose correctly in at least 10/12 trials for two consecutive sessions, they proceeded to the test trials.

Main test trials (sessions 1-4). The main test consisted of 16 trials (8 false belief and 8 true belief) administered over four sessions. To minimize potential confusion from switching back and forth between true and false belief trials, we administered each type of trial in blocks of four and only administered one type of trial in any given session. Also, at the beginning of every test session we gave the chimpanzees two warm-up trials in which we hid a grape (in one trial) and a banana (in the other trial) in the appropriate container and let the chimpanzee choose a container, as in the stage 2 training trials. Which reward was hidden first was determined randomly.

In the false belief trials, the Baiter showed the chimpanzee a grape or banana slice, put it into the yellow box, closed the box, and then placed the box onto the table and left the room, as in the final stages of training described above. However, in the false belief trials, the Switcher then approached the window, got the chimpanzee's attention, and switched the food inside the box for the other type of food. She did this in a sneaky manner by crouching down and hunching her shoulders, raising her eyebrows, making shushing sounds, and glancing repeatedly at the door to check if the Baiter was returning. When the Switcher was finished she sat down again and secretly signaled to the Baiter that she was finished by pressing a button on her walkie-talkie that caused the Baiter's walkie-talkie to beep. The Baiter then returned, positioned an occluder to block the chimpanzee's view of the containers, hid the box in a container, and removed the occluder. If chimpanzees understood that the Baiter had a false belief about the contents of the yellow box, they should choose the container corresponding to the box's *original* contents. The true belief procedure was identical to the false belief procedure except that the Switcher waited until the Baiter returned before switching the contents, the Baiter watched the switch, the switch was not sneaky, and the Baiter hid the box in the correct place according to its *current* contents.

Refresher session. For some chimpanzees, several weeks elapsed between the main test and the follow-up tests. We therefore administered another stage 3 training session just before the follow-up tests to ensure that chimpanzees remembered the correct location for each type of reward. Individuals had to get at least 10/12 trials correct in this session to proceed to the follow-up tests. All chimpanzees achieved this criterion.

Duping test (sessions 5 and 6). The two-session duping test investigated whether or not personally experiencing an unexpected change of contents would improve chimpanzees' performance in the false belief test. Three randomly chosen chimpanzees received the duping experience and the remaining two acted as non-duped controls. In the duping trials, the Baiter baited the yellow box as the chimpanzee observed, then put it down in the middle of the table and left the room. (The blue and white containers were not used in these trials and they were put aside.) The Switcher then pulled a curtain across the window to block the chimpanzee's view. She switched the food inside the box for the other type, put the box back onto the table, opened the curtain, returned to her seat, and secretly signaled to the Baiter that she was finished. The Baiter then returned, opened the box, and gave the chimpanzee the food, which was, of course, different from that the chimpanzee had seen placed into the box. Thus, chimpanzees were not required to make a choice in these trials; they merely observed as the experimenter opened the box and handed them the 'unexpected' food. While we could not be certain that the chimpanzees experienced surprise here in the same sense that a human would, a previous study has shown that chimpanzees do in fact notice when they encounter unexpected contents (Mendes, Rakoczy, & Call, 2008). In that study, chimpanzees, bonobos and gorillas were given the opportunity to continue searching in a box after they had already found one item inside. They did so significantly more often when the item they found was not the same as the one they had seen placed into the box moments before.

The non-duping control trials were identical to the duping trials except that the Switcher, behind the curtain, opened and closed the yellow box without switching its contents. The duped group received two duping trials and two non-duping trials in each session (alternating but always beginning with a non-duping trial). The non-duped control group received four non-duping trials in each session. For each group, these were followed immediately by four false belief trials in which chimpanzees had to choose between the blue and white containers, identical to the false belief trials in the main test described above. As in the main test, the correct response was for chimpanzees to choose the container normally associated with the food that the Baiter thought was in the yellow box (i.e., the original, rather than current, contents).

In-room test (session 7). In this final session, all chimpanzees received four false belief trials that were exactly like the false belief trials in the main test except that the Baiter remained in the room during the switch. After baiting the yellow box and placing it on the table, the Baiter stood up, turned around, and walked across the room, muttering to herself and pretending to be absorbed in fixing a panel on the opposite wall. The Switcher then quietly approached the window and, while the chimpanzee observed, sneakily switched the contents of the box, periodically checking to make sure the Baiter was not looking. After the Switcher was finished she sat down again and secretly signaled the Baiter by coughing inconspicuously. The Baiter then returned and carried out the remainder of the trial as before. Again, the correct choice was the container normally associated with the original contents of the yellow box, which the Baiter falsely believed to still be in there.

Coding and analyses

Chimpanzees' choice of container, defined as the first one they touched, was unambiguous and therefore coded live. We also coded from videotapes whether or not chimpanzees' spontaneous looking responses suggested apparent hesitation in choosing. For this measure, we noted whether or not chimpanzees looked back and forth between the two containers before making their choice. They were considered to have done so if they looked at the ultimately unchosen container *two or more* times before settling on the chosen container. This was somewhat different than in Krachun, Carpenter, et al.'s (2009) study, in which even one glance at the unchosen container was coded as a positive looking response. For the current study, we used the stricter criterion (looking at least twice, rather than at least once) because chimpanzees almost always looked at both containers at least once before choosing, making this behavior uninformative. We coded looking for all eight true belief and eight false belief trials of the main test only. Looking was not coded for the false belief trials of the follow-up tests because there were no corresponding true belief trials with which to compare them. The mean number of trials in which looking back and forth occurred was then compared across conditions. We also investigated looking in just a particular subset of these data, in keeping with Krachun, Carpenter, et al. (2009). In that study, true belief and false belief trials were examined on an equal footing by using only the subset of trials in which apes chose the same container as the competitor, as they had done throughout the pretest/training period. In keeping with this logic, in this study we also analyzed looking for just the trials in which chimpanzees chose the container corresponding to the current contents of the yellow box, as they had learned during training. Because each chimpanzee had a different number of data points for this analysis, we used proportion scores.

To determine inter-rater reliability, an independent coder who was blind to condition and naïve to the hypotheses of the study coded 25% of test trials for each measure. Reliability was excellent, with 100% agreement for choice and a Cohen's Kappa of 0.80 for looking. Statistical analyses for all measures were nonparametric and all *p* values reported below are exact and two-tailed.

Results and Discussion

In the training trials, the five chimpanzees took an average of 25 sessions to learn the unique locations for the two types of reward (range = 9 – 45 sessions). A sixth chimpanzee was still choosing randomly after 50 sessions and was therefore not tested further. There were no effects of order of condition on true belief or false belief performance (Mann-Whitney tests: $U \geq 0.50$, $N = 5$, $p \geq 0.30$ in both cases) and so we collapsed across order for all analyses. Because the main test and follow-up tests included varying numbers of false belief trials, we used proportion scores for easier comparison across tests.

Choice responses

We first examined chimpanzees' performance in the 16 main test trials (eight true belief and eight false belief trials; Sessions 1-4). As shown in Figure 5, chimpanzees succeeded in far more true belief trials than false belief trials, and the difference in group means approached significance (Wilcoxon test: $T^{+} = 15.00$, $N = 5$, $p = 0.063$). Chimpanzees' performance in false belief trials was not significantly different from the 50% chance probability of success ($T^{+} = 6.00$, N =3[2 ties], $p = 0.25$). In contrast, on average, chimpanzees got nearly 80% of true belief trials correct. While this is a high success rate it did not quite reach significance ($T^{+} = 15.00$, $N = 5$, $p = 0.063$), likely due to the small N. To investigate whether chimpanzees performed poorly in the false belief trials because they became confused by the repeated switching back and forth between the true and false belief conditions across test sessions, we also examined their performance in the first two trials only of each condition (i.e., the same number of test trials that children had received). However, the mean proportion of trials correct for this analysis (0.4 for false belief and 0.8 for true belief) was very similar to when we included all 16 trials of the main test. Furthermore, chimpanzees continued to perform poorly in the duping and in-room follow-up tests of Sessions 5 - 7 (see Figure 5).

Performance at the individual level was consistent with the group results. In the main test, all chimpanzees' true belief scores were higher than their false belief scores. Further, all true belief scores were above the chance value whereas all false belief scores were at or below chance (see Table 1). The follow-up manipulations resulted in no apparent improvements in chimpanzees' performance. In the duping test, the three chimpanzees who had been given personal experience with an unexpected change of contents did no better in the duping sessions than they had in the main test sessions (average proportion correct = 0.25 and 0.29, respectively; $T^+ = 1.00$, $N = 1$[2 ties], $p = 1.00$). Their performance was also not significantly different from that of the two non-duped chimpanzees, whose average proportion of trials correct in these follow-up sessions was 0.38 ($U = 1.50$, $N = 5$, $p = 0.50$). In the in-room test, chimpanzees' mean proportion of trials correct (0.15) was actually lower than in the main test (0.30), although this difference was again not significant ($T^+ = 12.00$, $N = 5$, $p = 0.25$).

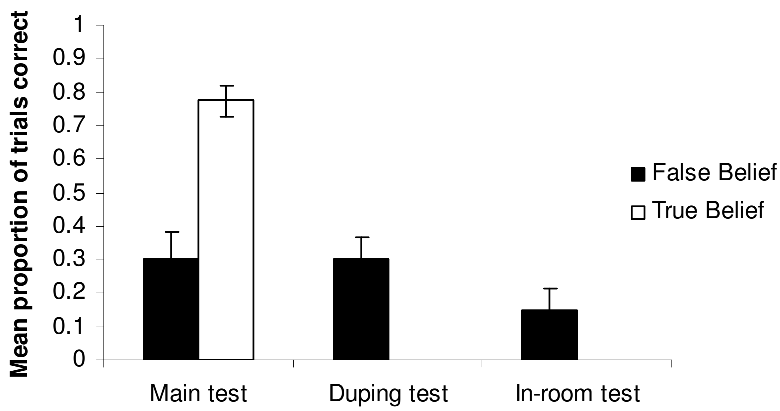

Figure 5. Mean (±*SE*) proportion of trials correct by chimpanzees in the main, duping and in-room tests of Study 2 (within-subjects, *N* = 5 for each bar). Note that no true belief trials were given in the duping and in-room tests.

Table 1
Proportion of correct true belief (TB) and false belief (FB) trials for individual chimpanzees in the main test and follow-up tests of Study 2

	Main test		Duping test	In-room test	All FB tests combined
Chimpanzee	(8 TB trials)	(8 FB trials)	(8 FB trials)	(4 FB trials)	(20 FB trials)
Alex	0.75	0.25	0.25[a]	0.00	0.20
Jahaga	0.88	0.50	0.38[a]	0.00	0.35
Fifi	0.75	0.13	0.13[a]	0.25	0.15
Trudy	0.63	0.50	0.25[b]	0.25	0.35
Annett	0.88	0.13	0.50[b]	0.25	0.30

[a] = duped chimpanzee

[b] = non-duped (control) chimpanzee

Because each chimpanzee received 20 false belief trials in total, including the main test and follow-up tests, it was possible to look for statistically significant effects in individual false belief performance using a binomial test. Two chimpanzees (Alex and Fifi) did significantly worse than chance (both *p*s < 0.01), while all other chimpanzees were at chance performance. Furthermore, the group mean for all 20 false belief trials (0.27) approached being significantly worse than chance ($T^{+} = 15.00$, $N = 5$, $p = 0.063$). In short, in their choice responses, chimpanzees gave no indication that they recognized the Baiter's false beliefs. The next question was whether their spontaneous looking responses suggested some degree of false belief understanding that was not apparent in their active choices.

Looking responses

The positive looking results found by Krachun, Carpenter, et al. (2009) were not replicated here: chimpanzees looked back and forth between containers in an equivalent proportion of true belief and false belief trials (true belief = 0.36; false belief = 0.33; $T^{+} = 4.00$, $N = 3$[2 ties], $p = 1.00$). This result did not change when we examined only the subset of true belief and false belief trials in which chimpanzees chose according to the current contents of the yellow box (true belief = 0.38; false belief = 0.33, $T^{+} = 10.00$, $N = 5$, $p = 0.63$). Thus, not only did chimpanzees fail to choose the correct container in our false belief test, they also failed to look at it more often in false belief trials than in true belief trials. Results for the looking measure, as for the choice measure, provided no evidence that chimpanzees recognized the Baiter's false belief about the contents of the yellow box on any level.

General Discussion

As in previous studies, chimpanzees did poorly in our new false belief test whereas human children showed a clear progression from below-chance performance at 3.5 years of age to above-chance performance at 4.5 years of age. However, in replicating the results of previous studies we have not simply succeeded in discovering once again what was already known. Our test differed from earlier tests in some important respects. Up until now all studies comparing false belief understanding in apes and children have relied exclusively on change-of-location procedures inspired by the standard verbal Sally-Anne test. We have expanded the repertoire of tasks available with a new false belief test for children and apes that focused not on the hidden location of an item, but on the hidden contents of a container, more along the lines of the standard verbal Smarties test. Our test was also more socially neutral to avoid the potentially problematic cooperative-communicative or competitive nature of previous tests. The fact that chimpanzees still failed our test suggests that the social context of previous tests was not what was causing chimpanzees to fail those tests. Instead, our findings support conclusions drawn from those previous studies that chimpanzees may be incapable of recognizing false belief states in others. This is further supported by the fact that chimpanzees failed our test despite all our attempts to make it easier for them, including giving them personal experience with being duped in the same way that the Baiter was duped. It is also supported by the fact that chimpanzees' spontaneous looking behavior did not reveal any signs of recognizing the Baiter's false beliefs on any level.

The obvious question raised by this and similar previous studies is why do chimpanzees do so poorly in false belief tests? There is some evidence to suggest that they are capable of recognizing some other types of mental states in others, including desires (Buttelmann, Call, & Tomasello, 2009a), intentions (Buttelmann, Carpenter, Call, & Tomasello, 2007), visual perceptions (Hare, Call, Agnetta, & Tomasello, 2000; Hare, Call, & Tomasello, 2006; Melis, Call, & Tomasello, 2006), and knowledge construed as what others have or have not seen in the recent past (Hare et al., 2001; Hirata & Matsuzawa, 2001; Menzel, 1974). Some have questioned the extent to which the methods used in those studies can distinguish responses based on recognizing mental states from those based on non-mentalistic mechanisms (e.g., Penn & Povinelli, 2007; Povinelli & Vonk, 2003, 2004). The fact remains, however, that chimpanzees passed those tests but they have never passed a false belief test. And so the question remains: what is it about false belief tests in particular that makes them so difficult for chimpanzees? We suggest two possibilities. The first, mentioned above, is that chimpanzees' failure to pass these tests may be indicative of a real inability to represent false belief states. This might in turn be due to the fact that they do not possess a complex natural language. A number of researchers argue that the human ability for language—in particular the mastery of sentential complements—is intimately connected with our ability to represent false belief states (e.g., de Villiers & Pyers, 2002). The claim is that being able to use and understand 'that' clauses (He thinks *that* there are Smarties in the box) may be necessary for even conceiving of situations in which people think things that are not the case in reality. Chimpanzees' consistent inability to pass false belief tests provides some support for this view. On the other hand, the recent studies reporting responses suggestive of false belief understanding in preverbal or barely verbal human infants (Buttelmann et al., 2009b; Onishi & Baillargeon, 2005; Surian et al., 2007) suggest that recognition of false beliefs on some level may be possible without language.

A second possibility is that chimpanzees may be capable of recognizing false beliefs, but that the tests created so far have been too difficult for them for other reasons. In the current study, for example, the version of the test developed for apes necessarily differed in some respects from the children's version. Most notably, the apes' test was entirely nonverbal while the children's test involved some verbal interaction. This is common in comparative research because using entirely nonverbal procedures with children, even when theoretically possible, would be too unnatural and inefficient. We were not overly concerned with this issue here because the approach has yielded positive results from the same and other apes in previous comparative studies on mental-state understanding (e.g., Call & Carpenter, 2001; Call & Tomasello, 2008; Kaminski et al., 2008, Study 1). However, some might argue that it could give children an advantage over apes in some types of procedures. For example, because our task could not be explained verbally to chimpanzees, they needed a large number of training trials to get them ready for the belief tests. The concern here is that they might have unthinkingly carried the strategy they learned during training trials over to the true and false belief test trials (i.e., choose based on the current contents of the yellow container). However, we do not think that this occurred because the chimpanzees almost always looked back and forth between the containers before making a choice, suggesting that they were not making an automatic, impulsive decision. Nevertheless, it would be desirable to find ways to minimize the training requirements in future false belief tests for chimpanzees.

It is also possible that the false belief tests devised so far for apes may ask too much of them in terms of attention and working memory. Despite our own and others' attempts to keep the tests as simple as possible, they have all required chimpanzees to attend closely to one or two experimenters as they moved rewards, containers, and themselves around in a sequence of events that may have been just too much for chimpanzees to keep track of. Perhaps what is needed is a simpler false belief test—one that is not too cognitively taxing for chimpanzees in other respects. While devising a simple nonverbal false belief test for chimpanzees has been a continuing challenge, a recent study by Krachun, Call, and Tomasello (2009) suggests a potentially fruitful approach. In that study, a visual illusion was used to investigate whether chimpanzees could recognize when their *own* mental states (specifically, their own visual perceptions) were false. Chimpanzees had to recognize that the appearance of food rewards displayed behind magnifying and minimizing lenses was perceptually misleading. Some chimpanzees passed the test by choosing the apparently smaller (but truly bigger) food reward significantly more often than chance, demonstrating that they recognized when they were experiencing a false perception. Chimpanzees may have performed better in that test than in false belief tests because the source of the false mental state was immediately perceptually apparent to them as they made their choice, which has not been the case in the false belief tests used to date. It would therefore be interesting to devise a social version of such a test to see if chimpanzees were also capable of recognizing when someone else was being perceptually misled by a visual illusion. Thus, just as some researchers have suggested that illusions could be useful in metacognition research (Allen & Bekoff, 1997; Carruthers, 2008), we suggest that they can also be an important tool for future comparative research into false belief understanding.

On a final note, it is interesting that while apes clearly did not do any better in our test than in previous ones, children did seem to find our test somewhat easier than those used in previous comparative studies (e.g., Call & Tomasello, 1999; O'Connell & Dunbar, 2003), as they passed it at a slightly younger age than reported in those studies. Beyond comparative research, our new paradigm is therefore a useful addition to the repertoire of tests available for testing human children with minimally verbal procedures that nevertheless require them to make active choice responses (rather than just looking responses). Researchers often administer two or more different types of false belief test to the same children in order to gain a more accurate picture of individual participants' abilities. The standard verbal Sally-Anne and Smarties tests are often paired together in the same study, for example. Such an aggregate approach to false belief testing is desirable, given that individual children do not always perform consistently across different false belief tests (e.g., Charman & Campbell, 1997; Holmes et al., 1996; Naito, 2003). Thus, our new change-of-contents test can complement existing procedures to allow for a more accurate assessment of false belief understanding, and its minimally verbal nature makes it particularly useful for testing children with low verbal skills.

References

Allen, C., & Bekoff, M. (1997). *Species of mind: The philosophy and biology of cognitive ethology*. Cambridge, MA: The MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46.

Buttelmann, D., Call, J., & Tomasello, M. (2009a). Do great apes use emotional expressions to infer desires? *Developmental Science, 12*, 688-698.

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science, 10,* F31-F38.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009b). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112,* 337-342.

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition, 3,* 207-220.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70*, 381-395.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12,* 187-192.

Carlson, S. M., Wong, A., Lemke, M., & Cosser, C. (2005). Gesture as a window on children's beginning understanding of false belief. *Child Development, 76,* 73-86.

Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology, 20*, 393-420.

Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language, 23,* 58-89.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two, three-, and four-year-olds' early theories of mind. *Child Development, 60,* 1263-1277.

Charman, T., & Campbell, A. (1997). Reliability of theory of mind task performance by individuals with a learning disability: A research note. *Journal of Child Psychology and Psychiatry, 38*, 725-730.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377-395.

De Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17,* 1037-1060.

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of three-year-olds' understanding of false belief. *Child Development, 62,* 83-97.

Hare, B. (2001). Can competitive paradigms increase the validity of experiments on primate social cognition? *Animal Cognition, 4*, 269-280.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour, 59,* 771-785.
Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*, 139-151.
Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition, 101,* 495-514.
Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative tasks. *Animal Behaviour, 68*, 571-581.
Hirata, S., & Matsuzawa, T. (2001). Tactics to obtain a hidden food item in chimpanzee pairs (*Pan troglodytes*). *Animal Cognition, 4,* 285-295.
Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*, 567-582.
Holmes, H. A., Black, C., & Miller, S. A. (1996). A cross-task comparison of false belief understanding in a head start population. *Journal of Experimental Child Psychology, 63*, 263-285.
Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition, 109,* 224-234.
Keenan, T., & Ellis, B. J. (2003). Children's performance on a false-belief task is impaired by activation of an evolutionarily-canalized response system. *Journal of Experimental Child Psychology, 85*, 236-256.
Krachun, C., Call, J., & Tomasello, M. (2009). Can chimpanzees distinguish appearance from reality? *Cognition,112,* 435-450.
Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science, 12,* 521-535.
Matsui, T., & Miura, Y. (2008). Pro-social motive promotes early understanding of false belief. Available from *Nature Precedings*. http://hdl.handle.net/10101/npre.2008.1695.1
Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology, 120,* 154-162.
Mendes, N., Rakoczy, H., & Call, J. (2008). Ape metaphysics: Object individuation without language. *Cognition*, *106*, 730-749.
Menzel, E. W. (1974). A group of young chimpanzees in a one-acre field. In A. Schrier & F. Stollnitz (Eds.), *Behavior of nonhuman primates* (Vol. 5, pp. 83-153). New York: Academic Press.
Naito, M. (2003). The relationship between theory of mind and episodic memory: Evidence for the development of autonoetic consciousness. *Journal of Experimental Child Psychology, 85*, 312-336.
O'Connell, S., & Dunbar, R. I. M. (2003). A test for comprehension of false belief in chimpanzees. *Evolution and Cognition, 9,* 131-140.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255-258.
Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences, 362,* 731-744.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125-137.

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: Suspiciously human? *Trends in Cognitive Sciences, 7,* 157-160.

Povinelli, D. J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind and Language, 19,* 1-28.

Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early deception and the child's theory of mind: False trains and genuine markers. *Child Development, 62,* 468-483.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*, 587-592.

Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56,* 135-148.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*, 580-586.

Washburn, D. A., Hopkins, W. D., & Rumbaugh, D. M. (1990). Effects of competition on video-task performance in monkeys (*Macaca mulatta*). *Journal of Comparative Psychology, 104,* 115-121.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*, 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.